



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12841

To link to this article : DOI :10.1007/978-3-319-10061-6_2
URL : http://dx.doi.org/10.1007/978-3-319-10061-6_2

To cite this version : Dégremont, Cédric and Venant, Antoine and Asher, Nicholas *[Semantic Similarity: Foundations](#)*. (2014) In: Logic and engineering of natural language semantics - LENLS10, 26 October 2013 - 28 October 2013 (Kanagawa, Japan).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Semantic Similarity: Foundations

Cédric Dégremont^(✉), Antoine Venant, and Nicholas Asher

IRIT, Toulouse, France

{cedric.degremont,antoine.venant,asher}@irit.fr

<http://www.irit.fr>

Abstract. This paper investigates measures of semantic similarity between conversations from an axiomatic perspective. We abstract away from real conversations, representing them as sequences of formulas, equipped with a notion of semantic interpretation that maps them into a different space. An example we use to illustrate our approach is the language of propositional logic with its classical semantics. We introduce and study a range of different candidate properties for metrics on such conversations, for the structure of the semantic space, and for the behavior of the interpretation function, and their interactions. We define four different metrics and explore their properties in this setting.

Keywords: Semantics · Distance · Metric · Similarity · Lattice · Conversations

1 Introduction

If linguistic behavior is to be analyzed as a form of rational behavior (Grice 1967), it is important to be able to assess the conversational goals of linguistic agents and the extent to which they are fulfilled by any given conversation in a manageable way. Specifying preferences over the set of all possible choices of what to say is clearly intractable for us as theorists and for speakers as practitioners. Instead, speakers must be able to group conversations into semantically similar classes and to assess the relative semantic proximity of any two pairs of conversations. The preferences of the agents over different ways of expressing themselves have to do with how close these ways are from satisfying certain positive or negative semantic goals. An elegant way to be able to do this, is to have a metric over conversations that is semantic in nature. The goal of this paper is to identify properties that characterize ‘semantic metrics’ and to identify reasonable axioms that can help us isolate well-behaved semantic metrics.

A workable definition of semantic distance between texts or conversations is also important for the evaluation of annotations of discourse structure in text and dialogue. It is also crucial to the success of the machine learning of semantic structures from annotated data, as all known algorithms rely on some

This research was supported by ERC Grant 269427.

notion of similarity or loss with respect to the target structure. While measures of syntactic similarity like ParseEval (Black et al. 1991) and Leaf Ancestor (Sampson 2000) are well-understood and used in computational linguistics, they yield intuitively wrong results. ParseEval, for instance, places too much importance on the boundaries of discourse constituents, which are often notoriously hard even for expert annotators to agree on. Investigations of distances between semantic interpretations of a text are rarely examined. While a natural equivalence and ordering relation over contents comes from the underlying logic of formal semantic analysis, this only gives a very crude measure. Some have appealed to a language of semantic primitives to exploit the more developed measures of syntactic distance in a more semantic setting. But such an approach depends on the choice of semantic primitives, with no clear consensus on how to go about determining these primitives.

Semantic distances are also relevant in the context of formal theories of belief revision. Lehmann et al. (2001) explores Alchourrón et al. (1985) style postulates that characterize a wide family of belief revision operator based on pseudo-distances on models satisfying only very mild assumptions. Our problem is also closely related to the problem of determining the distance of a scientific theory from the truth. This problem, referred to as the problem of verisimilitude or truthlikeness in philosophy of science (since Popper 1968), is arguably reducible to the problem of having a satisfactory concept of similarity between theories in a formal language.

The aim of this paper is to study semantic metrics for an abstract and simple concept of conversations. Syntactically, we assume that conversations are monoids with respect to concatenation. These conversations are equipped with an interpretation function mapping them into some distinct semantic space. In general, our assumptions about the semantic space and the interpretation function will be as minimal as possible. As far as identifying the axioms that characterize our concept of ‘semanticity’ for a metric goes, we will not be making any assumption. To analyze candidate axioms that characterize *well-behaved* semantic metrics, it will be interesting to consider the effect of assuming a bit more structure. Specifically we will pay some attention to the case in which the semantic co-domain of the interpretation function is a lattice. As an example, sequences of propositional formulas with their classical interpretation certainly fall under this category. We will moreover consider interpretation functions that satisfy some structural properties, for example assuming that the semantic meaning of a sequence is invariant under stuttering, that is immediate repetition of the same element in a sequence, or even assuming complete invariance under permutation.

To develop semantic metrics for conversations in natural language or for their representations in some formalism suitable for discourse interpretation (like for instance SDRT, Asher and Lascarides 2003) we first need to clarify the space of reasonable axioms and metrics for the simplest and most general representations. We take this first step here.

The paper is organized as follows. We start with a first section, Sect. 2, that contains technical preliminaries and settles the notation. We then describe in Sect. 3 different properties that can be met by an interpretation function, regarding how it interacts with sequences-concatenation or the structure of the semantic space. Section 4 introduces some elementary background about generalized metrics and metrics over subsets of a metric space. Section 5 draws a map of different level of *semanticity* for metrics, corresponding to different requirements on the interaction between the interpretation function and the metric. Section 6 introduces some concrete candidate semantic metrics. Sections 6–8 then describe how these potential measures of semantic similarity fare with respect to different lists of axioms. We show that certain combinations of axioms lead to trivialization results. We then conclude.

2 Preliminaries and Notation

This section contains some technical preliminaries and settles notation. The reader can skip this section on a first reading, and come back to it when needed.

2.1 Sets, Functions, Sequences, Orders and Lattices

Let X, Y be two sets. We let $X \ominus Y$ denote the symmetric difference of X and Y . Let $\text{card}(X)$ denote the cardinality of X . Let X^* be the set of finite strings over X . If $f : X \rightarrow Y$, we let $f(X)$ and $f[X]$ be alternative notation for the image of X under f , that is $f(X) = f[X] = \{f(x) | x \in X\}$. We let $\text{dom}(f) = X$, $\text{target}(f) = Y$ and $\text{ran}(f) = f(X)$. The kernel of f is the equivalence relation \sim , such that $x \sim y$ iff $f(x) = f(y)$. We say that f is *isotone* whenever for all x, y with $x \leq y$ we have $f(x) \leq f(y)$. We write $f : X \leftrightarrow Y$ whenever f is partial function from X to Y , that is there exists a non-empty subset of $A \subseteq Y$ such that $f : A \rightarrow Y$.

Given a sequence $\sigma \in X^*$ we let $\text{len}(\sigma)$ be the length of σ . If $k \leq \text{len}(\sigma)$ then we let $\sigma|_k$, be the prefix of σ of length k , and we let $\sigma(k)$ or $\sigma[k]$ be the k^{th} element of σ . We let $\text{ran}(\sigma)$ be the range of σ , that is $\text{ran}(\sigma) = \{\sigma[k] | 1 \leq k \leq \text{len}(\sigma)\}$. We let $\vec{\epsilon}$ be the empty sequence.

A relation \leq on X is a **pre-order** on X iff it is a reflexive and transitive relation on X . (X, \leq) is a **poset** iff \leq is a pre-order on X such that \leq is antisymmetric on X , that is $x \leq y$ and $y \leq x$, implies that $x = y$. A **lattice** (X, \leq) is a poset such that every two elements $x, y \in X$ have a least upper bound (or join, denoted $x \wedge y$) and a greatest lower bound (or meet, denoted $x \vee y$). A lattice (X, \leq) is bounded whenever X has a least and a greatest element (denoted \perp and \top).

2.2 Propositional Languages and Interpretation Functions

Given a language L we let $\varphi, \psi, \chi, \varphi_1, \varphi_2 \dots$ range over L , and we let $\vec{\varphi}, \vec{\psi}, \vec{\sigma}, \vec{\tau}, \vec{v}, \vec{\sigma_1}, \vec{\sigma_2} \dots$ range over L^* . Given a finite set $\text{PROP} = \{p_1, \dots, p_n\}$ we let $L_{\text{PROP}}(1)$ be defined as follows:

$$\varphi ::= p | \neg p | \top | \perp$$

where p ranges over PROP . And we define \mathbf{L}_{PROP} as follows:

$$\varphi ::= p \mid \top \mid \perp \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi$$

where p ranges over PROP . We define $\text{sig}(\varphi): \mathbf{L} \rightarrow \wp(\text{PROP})$ where $p \in \text{sig}(\varphi)$ iff p occurs in φ . Given a sequence $\vec{\sigma} \in \mathbf{L}^*$ or a subset $A \subseteq \mathbf{L}$, we write $\text{sig}(\vec{\sigma}) := \bigcup_{\psi \in \text{ran}(\vec{\sigma})} \text{sig}(\psi)$ or $\text{sig}(A) := \bigcup_{\psi \in A} \text{sig}(\psi)$, respectively.

Classical truth-functional interpretation of \mathbf{L}_{PROP} . Let $W_{\text{PROP}} = 2^{\text{PROP}}$. Depending on context, will treat a member $V \in W_{\text{PROP}}$ either as a function $V: \text{PROP} \rightarrow \{0, 1\}$ or as a subset $V \subseteq \text{PROP}$. These two representations are of course equivalent. We let $\llbracket \cdot \rrbracket^t: \mathbf{L} \rightarrow 2^{\text{PROP}}$, be the classical truth-functional interpretation function of $\mathbf{L}_{\text{PROP}}(1)$ and \mathbf{L}_{PROP} .

3 Properties of Interpretation Functions

In general, we will work with abstract concepts of a language and of an interpretation function. Let \mathbf{L}, \mathbf{X} be non-empty sets.

Definition 3.1 (Interpretation function). An interpretation function of \mathbf{L} into \mathbf{X} is a function $\llbracket \cdot \rrbracket: \mathbf{L}^* \rightarrow \mathbf{X}$.

An interpretation function for \mathbf{L} is an interpretation function \mathbf{L} into \mathbf{Y} for some non-empty set \mathbf{Y} .

3.1 Co-Domain

In this paper, we will sometimes assume that the semantic space has some structure. We are always explicit about these assumptions whenever we make them.

Definition 3.2. We say that $\llbracket \cdot \rrbracket$ is (W, \leq) -pre-order-valued ((W, \leq) -poset-valued) whenever $\text{target}(\llbracket \cdot \rrbracket) = W$ and (W, \leq) is a pre-order (respectively, a poset).

We say that $\llbracket \cdot \rrbracket$ is pre-order-valued, iff it is (W, \leq) -pre-order-valued for some pre-ordered set (W, \leq) , and similarly for poset-valued.

Definition 3.3. $\llbracket \cdot \rrbracket$ is (W, \leq, \wedge, \vee) -lattice-valued iff $\text{target}(\llbracket \cdot \rrbracket) = W$ and (W, \leq) is a lattice, with \wedge and \vee as its meet and join operator, respectively.

Definition 3.4. $\llbracket \cdot \rrbracket$ is $(W, \leq, \wedge, \vee, \perp, \top)$ -lattice-valued iff $\llbracket \cdot \rrbracket$ is (W, \leq, \wedge, \vee) -lattice-valued, and (W, \leq) is a bounded lattice, with \perp and \top as its least and greatest element, respectively.

We say that $\llbracket \cdot \rrbracket$ is lattice-valued, iff it is (W, \leq, \wedge, \vee) -lattice-valued for some (W, \leq) , \wedge and \vee . We say that $\llbracket \cdot \rrbracket$ is bounded lattice-valued, iff it is $(W, \leq, \wedge, \vee, \perp, \top)$ -lattice-valued, for some (W, \leq) , \wedge , \vee , \perp and \top .

Definition 3.5. $\|\cdot\|$ is set-valued iff we have $\text{target}(\|\cdot\|) = \wp(W)$ for some non-empty set W .

3.2 Structural Properties for Interpretation Functions

It will sometimes be interesting to restrict ourselves to interpretation functions satisfying certain structure properties. Assume that $\|\cdot\|$ is \leq -poset-valued. Below the comma ‘,’ is the concatenation operator. The axiom in the table below, are to be understood as quantifying universally. $\vec{\alpha}, \vec{\beta}$ ranging over $\text{dom}(\|\cdot\|)^*$ and α, β ranging over $\text{dom}(\|\cdot\|)$.

Axiom name	Meaning
contraction	$\ \vec{\alpha}, \varphi, \varphi, \vec{\beta}\ \leq \ \vec{\alpha}, \varphi, \vec{\beta}\ $
expansion	$\ \vec{\alpha}, \varphi, \vec{\beta}\ \leq \ \vec{\alpha}, \varphi, \varphi, \vec{\beta}\ $
exchange	$\ \vec{\alpha}, \varphi, \psi, \vec{\beta}\ = \ \vec{\alpha}, \psi, \varphi, \vec{\beta}\ $
right monotonicity	$\ \vec{\alpha}, \varphi\ \leq \ \vec{\alpha}\ $
left monotonicity	$\ \varphi, \vec{\alpha}\ \leq \ \vec{\alpha}\ $
$\vec{\epsilon} - \top$	$\ \vec{\alpha}\ \leq \ \vec{\epsilon}\ $
adjunction	If $\ \vec{\alpha}\ \leq \ \vec{\beta}\ $ and $\ \vec{\alpha}\ \leq \ \vec{\gamma}\ $ then $\ \vec{\alpha}\ \leq \ \vec{\beta} \vec{\gamma}\ $
mix	If $\ \vec{\alpha}_1\ \leq \ \vec{\beta}_1\ $ and $\ \vec{\alpha}_2\ \leq \ \vec{\beta}_2\ $ then $\ \vec{\alpha}_1 \vec{\alpha}_2\ \leq \ \vec{\beta}_1 \vec{\beta}_2\ $

For example, $\|\cdot\|$ satisfies contraction iff for every $\vec{\alpha}, \vec{\beta} \in L^*$ and $\varphi \in L$ we have $\|\vec{\alpha}, \varphi, \varphi, \vec{\beta}\| \leq \|\vec{\alpha}, \varphi, \vec{\beta}\|$.

Remark 3.6. If $\|\cdot\|$ satisfies exchange, then $\|\cdot\|$ for every $\vec{\varphi}$ and $\vec{\psi}$ that are equivalent up to permutation we have $\|\vec{\varphi}\| = \|\vec{\psi}\|$. If $\|\cdot\|$ satisfies either right or left monotonicity, then $\|\cdot\|$ satisfies $\vec{\epsilon} - \top$.

3.3 Stronger Properties

Definition 3.7 (Conjunctive, intersective interpretation)

- $\|\cdot\|$ is conjunctive iff it is lattice-valued and $\forall \vec{\varphi}, \vec{\psi}, \|\vec{\varphi} \vec{\psi}\| = \|\vec{\varphi}\| \wedge \|\vec{\psi}\|$.
- $\|\cdot\|$ is intersective iff it is set-valued and $\forall \vec{\varphi}, \vec{\psi}, \|\vec{\varphi} \vec{\psi}\| = \|\vec{\varphi}\| \cap \|\vec{\psi}\|$.

Definition 3.8. We let $\|\cdot\|^t$ be the interpretation function for L^* defined by $\|\vec{\varphi}\|^t := \llbracket \bigwedge_{\varphi \in \text{ran}(\vec{\varphi})} \varphi \rrbracket^t$.

Example 3.1. $\|\cdot\|^t$ is intersective. If $\|\cdot\|$ is intersective, then it is bounded \subseteq -lattice-valued.

4 Generalized Metrics

We start by recalling some basic definitions.

Definition 4.1 (Semi-Pseudometric). A semi-pseudometric on a set X is a function $d : (X \times X) \rightarrow \mathbb{R}$, such that for all $x, y, z \in X$ we have:

1. $d(x, x) = 0$;
2. $d(x, y) = d(y, x)$.

Definition 4.2 (Pseudometric). A pseudometric on a set X is a semi-pseudometric on X , such that for all $x, y, z \in X$ we have:

3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

If d is a (semi-)pseudometric on X , then (X, d) is a (semi-)pseudometric space.

Definition 4.3 (Trivial pseudo-metric). The trivial pseudo-metric over a set A is the function $d : \begin{cases} A \times A \rightarrow \mathbb{R} \\ \forall x, y \in A \quad d(x, y) = 0 \end{cases}$

4.1 Metrics on Valuations, Relations and Graphs

Given a finite set PROP Hamming distance on 2^{PROP} is the metric $\delta_{ham} : 2^{\text{PROP}} \times 2^{\text{PROP}} \rightarrow \omega$ defined as $\delta_{ham}(V, V') = \text{card}(V \ominus V')$.

4.2 Aggregators

Let δ_i be a pseudo-metric on a set X . We want to study closeness between subsets of X , and so we provide some natural aggregators α associating with δ_i a function $d_\alpha^i : 2^X \times 2^X \rightarrow \mathbb{R}$, that may or may not be a pseudo-metric, depending on the particular aggregator.

Definition 4.4 (min aggregator). Let $d_{\min}^i(A, B) = \min_{x \in A, y \in B} d_i(x, y)$.

Definition 4.5 (max aggregator). Let $d_{\max}^i(A, B) = \max_{x \in A, y \in B} d_i(x, y)$.

Definition 4.6 (Hausdorff aggregator). Formally $d_H^i(A, B) = \max\{\max_{x \in A} \min_{y \in B} d_i(x, y), \max_{y \in B} \min_{x \in A} d_i(x, y)\}$.

Definition 4.7 (mean aggregator). Formally $d_{am}^i(A, B) = \sum_{x \in A, y \in B} \frac{1}{\text{card}(A \times B)} d_i(x, y)$.

Remark 4.8. In general max and mean will return a non-zero value for (A, A) . Note also that the min aggregator will return 0 for (A, B) whenever $A \cap B \neq \emptyset$.

Let W be a set and let d be a pseudo-metric on W . Let \mathbb{L} be a language and let $\|\cdot\|$ be an interpretation function for L such that $\text{target}(\|\cdot\|) = \wp(W)$ for some non-empty set W . Let $d_\alpha^i : \wp(W) \times \wp(W) \rightarrow \mathbb{R}$ be an aggregator based on the distance d_i between points of W . We let $d_{\alpha, \mathbb{L}, \|\cdot\|}^i : \mathbb{L}^* \times \mathbb{L}^* \rightarrow \mathbb{R}$ be defined by $d_{\alpha, \mathbb{L}, \|\cdot\|}^i(\vec{\varphi}, \vec{\psi}) = d_\alpha^i(\|\vec{\varphi}\|, \|\vec{\psi}\|)$. When L and $\|\cdot\|$ are clear from context, we will simply write d_α^i for $d_{\alpha, \mathbb{L}, \|\cdot\|}^i$. For instance, $d_{H, \mathbb{L}^{prop}, \|\cdot\|}^{ham}$ is sometimes shortened as d_H^{ham} when \mathbb{L}^{prop} and $\|\cdot\|$ are clear from context.

5 What Is a Semantic Metric?

Now that we have set the stage for our investigations, our first task is to define our object of interest: semantic pseudometrics. Semantic pseudometrics are a subclass of linguistic pseudometrics.

Definition 5.1 (Linguistic (semi-)pseudometric). A linguistic (semi-)pseudometric on a language \mathbf{L} is a partial function $d : (\mathbf{L}^* \times \mathbf{L}^*) \rightarrow \mathbb{R}$ such that $\text{dom}(d)$ is symmetric and $(\text{dom}(d), d)$ is a (semi-)pseudometric space.

How semantic pseudometrics should be defined is not a fully straightforward matter. A minimal requirement would be the following:

$$\begin{aligned} &\text{If for every } \vec{\chi}_1, \vec{\chi}_2 \text{ with } \|\vec{\chi}_1\| = \|\vec{\chi}_2\| \\ &\text{we have } d(\vec{\varphi}, \vec{\chi}_1) = d(\vec{\psi}, \vec{\chi}_2) \text{ then } \|\vec{\varphi}\| = \|\vec{\psi}\| \quad (\text{min sem separation}) \end{aligned}$$

That is, if two sequences $\vec{\chi}_1, \vec{\chi}_2$ are semantically non-equivalent, then there should be two (other) semantically equivalent sequences, that are not pairwise equidistant from $\vec{\chi}_1$ and $\vec{\chi}_2$. A stronger, yet reasonable, assumption is:

$$\begin{aligned} &\text{If for every } \vec{\chi} \text{ we have } d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi}) \text{ then } \|\vec{\varphi}\| = \|\vec{\psi}\| \\ &\hspace{15em} (\text{sem separation}) \end{aligned}$$

The axiom states that if two sequences of formulas $\vec{\varphi}$ and $\vec{\psi}$ are not semantically equivalent, then there is some sequence of formulas $\vec{\chi}$ that is not at the same distance from both $\vec{\varphi}$ and $\vec{\psi}$.

Fact 5.2. *Let d be a semi-pseudometric. If d satisfies (sem separation), then it satisfies (min sem separation).*

Finally we consider a stronger axiom:

$$\begin{aligned} &\text{If } d(\vec{\varphi}, \vec{\psi}) = 0 \text{ then } \|\vec{\varphi}\| = \|\vec{\psi}\| \hspace{10em} (\text{zero} \Rightarrow \text{sem} \equiv) \end{aligned}$$

The axiom is a regularity condition stating, that semantically non-equivalent sequences of formulas, should be at positive distance of each other.

Fact 5.3. *Let d be a semi-pseudometric. If d satisfies (zero \Rightarrow sem \equiv), then it satisfies (sem separation).*

The two become equivalent if we assume triangle inequality.

Fact 5.4. *Let d be a pseudo-metric. d satisfies (sem separation) iff d satisfies (zero \Rightarrow sem \equiv).*

The converse of (zero \Rightarrow sem \equiv), below, states that semantically equivalent sequences formulas, should be a distance 0 of each other.

$$\begin{aligned} &\text{If } \|\vec{\varphi}\| = \|\vec{\psi}\| \text{ then } d(\vec{\varphi}, \vec{\psi}) = 0 \hspace{10em} (\text{sem} \equiv \Rightarrow \text{zero}) \end{aligned}$$

Unsurprisingly (sem $\equiv \Rightarrow$ zero) will filter out syntactically driven notions such as δ_{count} or $\delta_{\text{synt, count}}$.

Definition 5.5. Given a language L , let $\delta_{count}(\vec{\varphi}, \vec{\psi}) := \text{card}(\text{ran}(\vec{\varphi}) \ominus \text{ran}(\vec{\psi}))$

Definition 5.6. Given a language L , let $\delta_{synt,count}(\vec{\varphi}, \vec{\psi}) := \delta_{count}(\vec{\varphi}, \vec{\psi}) + \text{card}(\text{sig}(\vec{\varphi}) \ominus \text{sig}(\vec{\psi}))$

Fact 5.7. $(L_{\text{PROP}}(1), \delta_{count})$ does not satisfy $(\text{sem} \equiv \Rightarrow \text{zero})$.

Fact 5.8. $(L_{\text{PROP}}(1), \delta_{synt,count})$ does not satisfy $(\text{sem} \equiv \Rightarrow \text{zero})$.

As observed previously, $(\text{sem} \equiv \Rightarrow \text{zero})$ rules out a number of aggregators, e.g. :

Fact 5.9. $(L_{\text{PROP}}, d_{\text{max}}^{\text{ham}})$ does not satisfy $(\text{sem} \equiv \Rightarrow \text{zero})$.

We have seen that $(\text{sem} \equiv \Rightarrow \text{zero})$ and the triangle inequality together imply that two semantically equivalent points are equidistant to any other third point. This latter notion of semantic invariance implies in return $(\text{sem} \equiv \Rightarrow \text{zero})$ and might be a desirable property as well:

$$\text{If } \|\vec{\varphi}\| = \|\vec{\psi}\| \text{ then for every } \vec{\chi} \text{ we have } d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi}) \quad (\text{sem preservation})$$

Fact 5.10. Let d be a semi-pseudometric. If d satisfies (sem preservation) then d satisfies $(\text{sem} \equiv \Rightarrow \text{zero})$.

Finally, we can require our (semi-)pseudometric to be fully induced by a distance on the co-domain of the interpretation function $\|\cdot\|$, which we define as follows:

$$\text{If } \|\vec{\varphi}_1\| = \|\vec{\psi}_1\| \text{ and } \|\vec{\varphi}_2\| = \|\vec{\psi}_2\| \text{ then } d(\vec{\varphi}_1, \vec{\varphi}_2) = d(\vec{\psi}_1, \vec{\psi}_2) \quad (\text{sem induced})$$

Fact 5.11. Let d be a semi-pseudometric. If d satisfies (sem induced) then d satisfies (sem preservation).

Fact 5.12. Let d be a pseudo-metric. If d satisfies $(\text{sem} \equiv \Rightarrow \text{zero})$, then it satisfies (sem induced).

Corollary 5.13. Let d be a pseudo-metric that satisfies $(\text{sem} \equiv \Rightarrow \text{zero})$. There exists a pseudo-metric \dot{d} on $\text{ran}(\|\cdot\|)$ such that $\dot{d}(\|\vec{\varphi}\|, \|\vec{\psi}\|) = d(\vec{\varphi}, \vec{\psi})$.

Fact 5.14. Let d be a pseudo-metric that verifies $(\text{sem} \equiv \Rightarrow \text{zero})$. Let \equiv be the kernel of $\|\cdot\|$. The following holds:

$$1. \text{ If } \vec{\varphi} \equiv \vec{\psi} \text{ then } \forall \vec{\chi} \ d(\vec{\chi}, \vec{\varphi}) = d(\vec{\chi}, \vec{\psi}).$$

Fact 5.15. Let d be a pseudo-metric. If d satisfies (min sem separation) and $(\text{sem} \equiv \Rightarrow \text{zero})$, then it satisfies (sem separation).

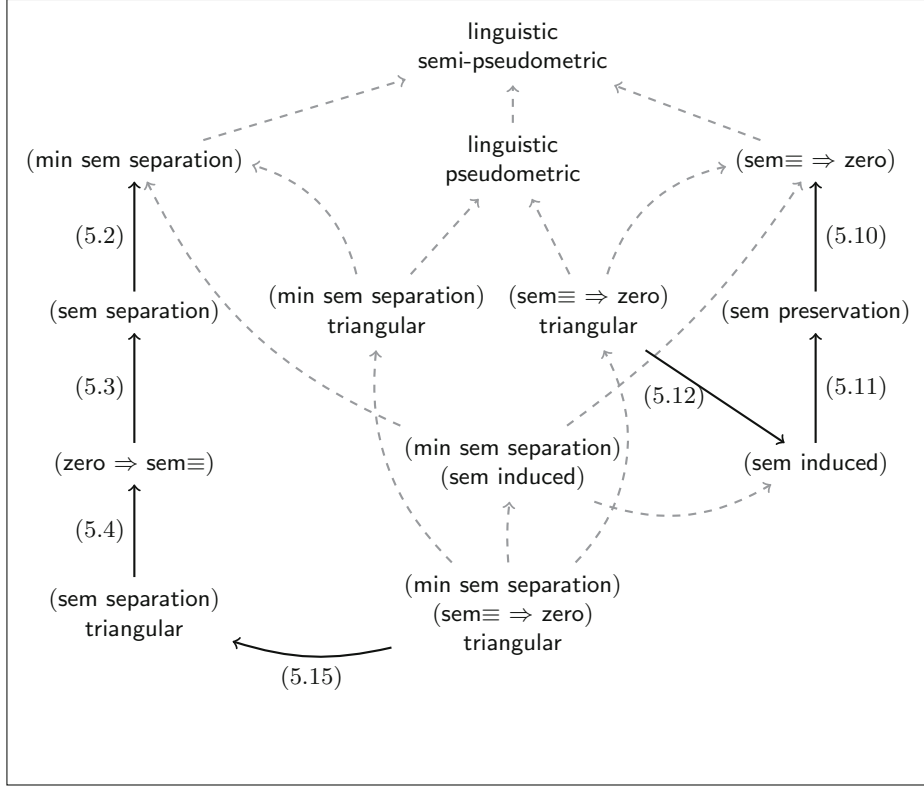


Fig. 1. Summary of the results in Sect. 5. Dashed arrows follow from definitions.

Figure 1, in p. 9, summarizes the relation between the axioms discussed in this section. We are now ready to define our notion of ‘semanticity’.

Definition 5.16 (Semantic Pseudometric). A linguistic (semi-)pseudometric is *semantic* whenever for all $\vec{\varphi}, \vec{\psi}$ we have $\|\vec{\varphi}\| = \|\vec{\psi}\|$ iff for all $\vec{\chi}_1, \vec{\chi}_2$ such that $\|\vec{\chi}_1\| = \|\vec{\chi}_2\|$ we have $d(\vec{\varphi}, \vec{\chi}_1) = d(\vec{\psi}, \vec{\chi}_2)$.

Fact 5.17. A linguistic semi-pseudometric is *semantic* iff it satisfies (min sem separation) and (sem induced).

Fact 5.18. A linguistic pseudometric is *semantic* iff it satisfies (min sem separation) and (sem $\equiv \Rightarrow$ zero).

Now that we have settled our definition of semantic pseudo metric, which we will use in the sequel, we can tackle our main problem—in brief:

What are reasonable properties of a semantic pseudometric on (a subset of) L^* ?

6 Axioms for Semantic Metrics

In this section we will investigate how various semantic pseudo metrics behave with respect to some different axioms, some more intuitive than others. We have already introduced the Hausdorff metric d_H^{ham} over sequences of formulas in \mathbf{L}_{PROP} . We now provide some other metrics.

6.1 Examples of Semantic Metrics

If $\|\cdot\|$ is finite set-valued, we can define the following metrics:

Definition 6.1 (Semantic Symmetric Difference Metric). The semantic symmetric difference metric is the cardinality of the symmetric difference between the respective interpretations. $d_{\ominus}(\vec{\varphi}, \vec{\psi}) := \text{card}(\|\vec{\varphi}\| \ominus \|\vec{\psi}\|)$.

Definition 6.2 (Proportional metric). The proportional metric decreases from 1 to 0 as the ratio between the intersection and the union of the respective interpretations increases.

$$d_{\alpha}(\vec{\varphi}, \vec{\psi}) = \begin{cases} 0 & \text{if } \|\vec{\varphi}\| = \|\vec{\psi}\| = \emptyset \\ 1 - \frac{\text{card}(\|\vec{\varphi}\| \cap \|\vec{\psi}\|)}{\text{card}(\|\vec{\varphi}\| \cup \|\vec{\psi}\|)} & \text{otherwise} \end{cases}$$

Observe that, with this measure, pairs of sequences with non-empty disjoint interpretations will always be at distance 0 of each other.

The idea is behind the next metric is to measure closeness by the continuations of two sequences. Our idea is to count how many possible *semantic continuations* one sequence has that can not apply to the other. For a set-valued interpretation function, this is given by the cardinal of the symmetric difference between the power sets of the respective interpretations. This should then be normalized by the possible ways to continue these sequences. This is given by the cardinal of the cartesian product of the power sets of the respective interpretations. Moreover we normalize this measure by the number of possible semantic values, $\text{card}(\wp(\{0, 1\}^{PROP})) = 2^{2^{\text{card}(PROP)}}$. Here is the definition:

Definition 6.3 (Continuation-based pseudometric)

$$d_C(\vec{\varphi}, \vec{\psi}) = \frac{\text{card}(2^{\|\vec{\varphi}\|} \ominus 2^{\|\vec{\psi}\|})}{\text{card}(2^{\|\vec{\varphi}\|} \times 2^{\|\vec{\psi}\|})} \cdot \frac{1}{2^{2^{\text{card}(PROP)}}}$$

In the simplistic setting of building a semantic metric for \mathbf{L}_{PROP}^* with the classical truth functional $\|\cdot\|^t$ interpretation, a natural notion of understanding *possible continuation*, is one of *consistent continuation*. A set $\|\vec{\varphi}\|^t$ is a possible continuation of $\vec{\psi}$ just in case $\|\vec{\varphi}\|^t \cap \|\vec{\psi}\|^t \neq \emptyset$. Following this definition, a possible continuation of $\vec{\varphi}$ which is not a possible continuation of $\vec{\psi}$ is uniquely decomposed into a part of $\|\vec{\varphi}\|^t \setminus \|\vec{\psi}\|^t$ and a set of valuations that are neither in $\|\vec{\varphi}\|^t$ nor $\|\vec{\psi}\|^t$.

Fact 6.4. *For any set-valued interpretation, the preceding definition is equivalent to the following:*

$$d_C(\vec{\varphi}, \vec{\psi}) = \frac{2^{\text{card}(\|\vec{\varphi}\|)} + 2^{\text{card}(\|\vec{\psi}\|)} - 2 \cdot 2^{\text{card}(\|\vec{\varphi}\| \cap \|\vec{\psi}\|)}}{2^{\text{card}(\|\vec{\varphi}\|) + \text{card}(\|\vec{\psi}\|)}}$$

6.2 Shortest Paths in Covering Graphs

Monjardet (1981) summarizes interesting results concerning metrics on posets and lattices. We will make use of two of these results to shed a different light on the semantic metrics defined in the previous section. In what follows, let $\langle W, \vee, \top, \leq \rangle$ be a bounded semi-lattice with \top as greatest element. For all $x, y \in W$, we say that y covers x iff $x < y$ and $\forall z, x < z \leq y \rightarrow y = z$. Define also inductively the *rank* of an element, by setting all $<$ -minimal elements of rank 0, and for every y covering a x of rank n , setting y of rank $n + 1$.

Definition 6.5 (Covering graph of a semi-lattice). Let $\langle W, \top, \leq \rangle$ be a semi-lattice with \top as greatest element. The **covering graph** $G(W) = \langle V, E \rangle$ of W is such that $V = W$ and $(x, y) \in E$ iff x covers y or y covers x .

Definition 6.6. An **upper valuation** is an isotone map $v : W \rightarrow \mathbb{R}$ such that $\forall z, z \leq x, y \rightarrow v(x) + v(y) \geq v(x \vee y) + v(z)$.

Let $G(W) = \langle W, E \rangle$ be the covering graph of W and let v be an upper valuation on W . For each edge $(x, y) \in E$, let the weight function $\omega_v : E \rightarrow \mathbb{R}$ induced by v be defined by $\omega_v(x, y) = |v(x) - v(y)|$. Moreover, let $\pi(x, y)$ be the set of paths from x to y . We make use of two results exposed in Monjardet (1981):

Fact 6.7 (Monjardet 1981). *Let v be an isotone upper valuation. We have:*

1. *the function $d_v(x, y) = 2v(x \vee y) - v(x) - v(y)$ is positive and verifies the triangle inequality.*
2. $d_v(x, y) = \delta_v(x, y) := \min_{p_{xy} \in \pi(x, y)} \sum_{(z_1, z_2) \in p_{xy}} \omega_v(x, y).$

Definition 6.8. If v is an isotone, positive, upper valuation that assigns 0 to minimal elements in the semi-lattice, then the **normalized distance** is defined by

$$d_v^n(x, y) = \begin{cases} \frac{d_v(x, y)}{v(x \vee y)} & \text{if } v(x \vee y) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Fact 6.9. *If v is an isotone, positive, upper valuation that assigns 0 to minimal elements in the semi-lattice, then the normalized distance $d_v^n \geq 0$ and d_v^n verifies the triangle inequality.*

This offers a new perspective on the Symmetric difference and Proportional metrics as metrics defined by minimal-weighted paths in the lattice:

Fact 6.10. Consider the (semi-)lattice $\langle 2^{prop}, \subseteq, \text{PROP} \rangle$, and the mapping v_0 that assigns to each $V \in 2^{prop}$ its rank. We have

$$d_{\ominus}(\vec{\varphi}, \vec{\psi}) = \delta_{v_0}(\|\vec{\varphi}\|, \|\vec{\psi}\|) = \text{card}(\|\vec{\varphi}\| \ominus \|\vec{\psi}\|)$$

Fact 6.11. Consider the (semi-)lattice $\langle 2^{prop}, \subseteq, \text{PROP} \rangle$, and the mapping v_0 that assigns to each $V \in 2^{prop}$ its rank. We have

$$d_{\alpha}(\vec{\varphi}, \vec{\psi}) = \delta_{v_0}^n(\|\vec{\varphi}\|, \|\vec{\psi}\|)$$

When the lattice is the lattice of subsets of some set A , the rank of $X \subseteq A$ coincide with its cardinal, hence the two facts above. This suggest d_{v_0} and $d_{v_0}^n$ as natural generalisations of d_{\ominus} and d_{α} for more general semi-lattices.

The continuations-based metric is also expressible in these minimal-weighted paths terms, but this requires a little more work:

Fact 6.12. Let $v : W \rightarrow \mathbb{R}$ be an isotone, mapping. The mapping $w : W \rightarrow \mathbb{R}$ such that $w : x \mapsto -2^{-v(x)}$ is isotone as well. Moreover, if v is an upper valuation, then so is w .

Corollary 6.13. Consider the (semi-)lattice $\langle 2^{prop}, \subseteq, \text{PROP} \rangle$, and the mapping v_0 that assigns to each $V \in 2^{prop}$ its rank, and w_0 defined by $w_0(x) = -2^{-v_0(x)}$. We have

$$d_{\mathcal{C}}(\vec{\varphi}, \vec{\psi}) = \delta_{w_0}(\|\vec{\varphi}\|, \|\vec{\psi}\|)$$

Corollary 6.14. d_{\ominus}, d_{α} and $d_{\mathcal{C}}$ are all pseudo-metrics.

6.3 Stronger Semantic Axioms

We next move to axioms that differentiate between our metrics. We start by considering the following axiom:

$$d(\vec{\varphi}, \vec{\varphi} \vec{\psi}) \leq d(\vec{\varphi}, \vec{\psi}) \quad (\text{rebar property})$$

From the four metrics we have introduced, only d_H^{ham} does not satisfy it.

Fact 6.15. Let $\|\cdot\|$ be intersective. d_{\ominus}, d_{α} and $d_{\mathcal{C}}$ satisfy (rebar property).

Fact 6.16. Let $\|\cdot\| = \|\cdot\|^t$. $(L_{\text{PROP}}, d_H^{ham})$ does not satisfy (rebar property).

Assume that $\|\cdot\|$ is \preceq -poset-valued. The following axiom is very mild:

$$\text{If } \|\vec{\varphi}\| \preceq \|\vec{\psi}\| \text{ then } d(\vec{\varphi}, \vec{\varphi} \vec{\psi}) \leq d(\vec{\varphi}, \vec{\psi}) \quad (\text{antitonicity})$$

Fact 6.17. Let $\|\cdot\|$ satisfy adjunction and right weakening and let d be a semi-pseudometric. If d satisfies (sem $\equiv \Rightarrow$ zero), then d satisfies (antitonicity).

6.4 Domain Axioms for Semantic Pseudometrics

Domain axioms require the metric to be well-defined on large portions of the language. Given an interpretation function $\|\cdot\| : L^* \rightarrow D$ of these sequences of sentences into some co-domain, we could expect to have:

$$\text{If } \vec{\varphi}, \vec{\psi} \in L^*, \text{ then } d(\vec{\varphi}, \vec{\psi}) \in \mathbb{R}. \quad (\text{linguistic domain})$$

It should be realized that the preceding axiom is relatively strong. Consider for example d_H^{ham} .

Fact 6.18. *But $(L_{\text{PROP}}, d_H^{ham})$ does not satisfy (linguistic domain).*

Fact 6.19. *d_\ominus, d_α, d_C verify (linguistic domain)*

Weakening the preceding axiom, without dropping it entirely, can be done if the co-domain of $\|\cdot\|$ is a bounded poset. Recall that \perp denote the least element of a bounded poset.

$$\text{If } \|\vec{\varphi}\| \neq \perp \text{ and } \|\vec{\psi}\| \neq \perp, \text{ then } d(\vec{\varphi}, \vec{\psi}) \in \mathbb{R}. \quad (\text{consistent domain})$$

Fact 6.20. *$(L_{\text{PROP}}, d_H^{ham})$ satisfies (consistent domain).*

6.5 Axioms for Set-Valued Semantic Pseudometrics

If $\|\cdot\|$ is set-valued and $\text{target}(\|\cdot\|) = \wp(W)$ for some W and (W, δ) is a metric space then we can investigate axioms like the following one considered in Eiter and Mannila (1997):

$$\text{whenever } \|\vec{\varphi}\| = \{w\} \text{ and } \|\vec{\psi}\| = \{v\} \text{ then } d(\vec{\varphi}, \vec{\psi}) = \delta(w, v) \quad (\text{EM})$$

A semantic metric defined as an aggregator of the values of the distance between points in the interpretation of either sequences, will satisfy the preceding axiom.

6.6 Signature Invariance Axioms

The next condition states that the relative proximity of conversations should not depend on irrelevant aspects pertaining to the choice of signature.

$$\begin{aligned} &\text{If } \vec{\varphi}, \vec{\psi}, \vec{\chi} \in L'^* \text{ and } L' \subseteq L, \text{ then we have} \quad (\text{weak sig inv}) \\ &d_L(\vec{\varphi}, \vec{\chi}) \leq d_L(\vec{\psi}, \vec{\chi}) \text{ iff } d_{L'}(\vec{\varphi}, \vec{\chi}) \leq d_{L'}(\vec{\psi}, \vec{\chi}) \end{aligned}$$

Fact 6.21. *$(L_{\text{PROP}}(1), \delta_{\text{count}})$ and $(L_{\text{PROP}}(1), \delta_{\text{synt}, \text{count}})$ satisfy (weak sig inv).*

Fact 6.22. *$d_H^{ham}, d_\ominus, d_\alpha, d_C$ satisfy (weak sig inv)*

7 Preservation Axioms

7.1 Uniform Preservation Axioms

The following axiom states that extending conversations with a given piece of information should not change the relative proximity of conversations. Formally:

$$\text{If } d(\vec{\varphi}, \vec{\chi}) \leq d(\vec{\psi}, \vec{\chi}) \text{ then } d(\vec{\varphi}\vec{\varphi}_0, \vec{\chi}\vec{\varphi}_0) \leq d(\vec{\psi}\vec{\varphi}_0, \vec{\chi}\vec{\varphi}_0) \quad (\text{uniform preservation})$$

But such an axiom can lead to triviality.

Fact 7.1. *Let $\|\cdot\|$ satisfy exchange, contraction and expansion and let d be a pseudometric. If d satisfies (uniform preservation) and $(\text{sem} \equiv \Rightarrow \text{zero})$, then whenever $d(\varphi, \chi) \leq d(\psi, \chi)$ then $d(\psi\chi, \varphi\psi\chi) = 0$.*

The next corollary is slightly technical. Let us introduce a bit of notation. Let $o : L^* \rightarrow \mathbb{R}$ be defined by $o(\vec{\varphi}) = d(\vec{\epsilon}, \vec{\varphi})$, and let \sim be the kernel of o . Let \leq_o be the total pre-order induced by o , with $\vec{\varphi} \leq_o \vec{\psi}$ iff $o(\vec{\varphi}) \leq o(\vec{\psi})$. Let $[\vec{\varphi}]$ be the equivalence class of $\vec{\varphi}$ in L^* / \sim . Let $\downarrow[\vec{\varphi}] = \{\vec{\psi} \in L^* \mid o(\vec{\psi}) \leq o(\vec{\varphi})\}$ and let $\downarrow[\vec{\varphi}]^*$ be the reflexive transitive closure of $\downarrow[\vec{\varphi}]$.

Corollary 7.2. *Let $\|\cdot\|$ satisfy exchange, contraction, expansion and $\vec{\epsilon} - \top$ and let d be a pseudometric. If d satisfies (uniform preservation) and $(\text{sem} \equiv \Rightarrow \text{zero})$, then for every $\vec{\varphi}, \vec{\chi} \in [\vec{\varphi}]$, $\vec{\psi}_1, \vec{\psi}_2 \in \downarrow[\vec{\varphi}]^*$ we have $d(\vec{\varphi}, \vec{\psi}_1\vec{\chi}\vec{\psi}_2) = 0$.*

Corollary 7.3. *Let $\|\cdot\|$ satisfy exchange, contraction, expansion and $\vec{\epsilon} - \top$ and let d be a pseudometric. If d satisfies (uniform preservation), $(\text{sem} \equiv \Rightarrow \text{zero})$ and (min sem separation), then for every $\vec{\varphi}, \vec{\chi} \in [\vec{\varphi}]$, $\vec{\psi}_1, \vec{\psi}_2 \in \downarrow[\vec{\varphi}]^*$ we have $\|\vec{\varphi}\| = \|\vec{\psi}_1\vec{\chi}\vec{\psi}_2\|$.*

Hence (uniform preservation) comes with very disputable consequences. The converse is even more problematic:

$$\text{If } d(\vec{\varphi}\vec{\varphi}_0, \vec{\chi}\vec{\varphi}_0) \leq d(\vec{\psi}\vec{\varphi}_0, \vec{\chi}\vec{\varphi}_0) \text{ then } d(\vec{\varphi}, \vec{\chi}) \leq d(\vec{\psi}, \vec{\chi}) \quad (\text{uniform anti-preservation})$$

These two axioms are quite demanding.

Fact 7.4. *Let $\|\cdot\| = \|\cdot\|^t$. d_H^{ham} satisfy neither (uniform preservation), nor uniform anti-preservation.*

Fact 7.5. *Let $\|\cdot\| = \|\cdot\|^t$. d_α , d_\ominus and d_C satisfy neither (uniform preservation), nor (uniform anti-preservation).*

But the situation is much more radical for (uniform anti-preservation): if the interpretation satisfies very mild conditions: such as contraction, expansion and exchange, then the only semi-pseudometric satisfying (uniform anti-preservation) and $(\text{sem} \equiv \Rightarrow \text{zero})$, is the trivial metric.

Fact 7.6. *Let $\|\cdot\|$ be an interpretation satisfying contraction, expansion and exchange and let d be a semi-pseudometric on $\text{target}(\|\cdot\|)$. The following are equivalent:*

1. *d satisfies (uniform anti-preservation) and $(\text{sem} \equiv \Rightarrow \text{zero})$*
2. *d is the trivial metric on $\text{target}(\|\cdot\|)$*

This result is a very strong argument against the reasonableness of (uniform anti-preservation).

7.2 Preservation Axioms: Close Information

The preceding axioms considered extensions of two sequences with the same sequence of formulas. As we have seen, they are too demanding. What if instead, we are interested in the relative effect of extending with a sequences that might be more or less similar to the sequence it is extending. We could expect, that the closer that new sequence is from the original one, the closer the resulting conversation will be from the original one. Or at least that a reverse in respective orderings cannot occur. Formally,

$$\text{If } d(\vec{\varphi}, \psi_1) < d(\vec{\varphi}, \psi_2) \text{ then } d(\vec{\varphi}, \vec{\varphi}\psi_1) \leq d(\vec{\varphi}, \vec{\varphi}\psi_2) \quad (\text{action pref})$$

Fact 7.7. *$d_H^{\text{ham}}, d_\ominus, d_\alpha, d_C$ do not satisfy (action pref)*

Conversely, we can require the deviation of the resulting sequence to be smaller, whenever the original sequence is closer to the new one by which it is extended.

$$\text{If } d(\vec{\varphi}, \chi) < d(\vec{\psi}, \chi) \text{ then } d(\vec{\varphi}, \vec{\varphi}\chi) < d(\vec{\psi}, \vec{\psi}\chi) \quad (\text{coherent deviation})$$

Fact 7.8. *Let $\|\cdot\| = \|\cdot\|^t$. d_α satisfy neither (action pref) nor (coherent deviation).*

Fact 7.9. *Let d be a semi-pseudometric on $\text{target}(\|\cdot\|)$. The following are equivalent:*

1. *d satisfies (coherent deviation) and the triangle inequality*
2. *d is the trivial metric on $\text{target}(\|\cdot\|)$*

As we will show, the respective converses of the two preceding axioms are certainly unreasonable.

$$\text{If } d(\vec{\varphi}, \vec{\varphi}\psi_1) \leq d(\vec{\varphi}, \vec{\varphi}\psi_2) \text{ then } d(\vec{\varphi}, \psi_1) \leq d(\vec{\varphi}, \psi_2) \quad (\text{converse strong action pref})$$

Fact 7.10. *Let d be a pseudometric on $\text{target}(\|\cdot\|)$. The following are equivalent:*

1. *d satisfies (converse strong action pref)*
2. *d is the trivial metric on $\text{target}(\|\cdot\|)$*

If $d(\vec{\varphi}, \vec{\varphi}_\chi) \leq d(\vec{\psi}, \vec{\psi}_\chi)$ then $d(\vec{\varphi}, \chi) \leq d(\vec{\psi}, \chi)$
 (converse coherent deviation)

Fact 7.11. *Let d be a pseudometric on $\text{target}(\|\cdot\|)$. The following are equivalent:*

1. d satisfies (converse coherent deviation)
2. d is the trivial metric on $\text{target}(\|\cdot\|)$.

8 Conjunction and Disjunction Axioms

8.1 Conjunction Axioms

Assume that $\|\cdot\|$ is lattice-valued. The following axiom regulates the behavior of the distance with respect to the meet. But as we will see, it is much too demanding.

If $\|\vec{\varphi}_1\| \wedge \|\vec{\varphi}_2\| \leq \|\vec{\varphi}_1\| \wedge \|\vec{\varphi}_3\|$ then $\delta(\vec{\varphi}_1, \vec{\varphi}_2) \geq \delta(\vec{\varphi}_1, \vec{\varphi}_3)$ (strong \wedge rule)

Fact 8.1. *Let $\|\cdot\|$ be lattice-valued and let d be a semi-pseudometric on $\text{target}(\|\cdot\|)$. If d satisfies (strong \wedge rule), then whenever $\|\vec{\varphi}\| \leq \|\vec{\psi}\|$, we have $d(\vec{\chi}, \vec{\varphi}) \geq d(\vec{\chi}, \vec{\psi})$ for any $\vec{\chi}$.*

Corollary 8.2. *Let $\|\cdot\|$ be lattice-valued and let d be a semi-pseudometric on $\text{target}(\|\cdot\|)$. If d satisfies (strong \wedge rule), then whenever $\|\vec{\varphi}\| \leq \|\vec{\psi}\|$, we have $d(\vec{\varphi}, \vec{\psi}) = d(\vec{\psi}, \vec{\varphi}) = 0$.*

Corollary 8.3. *Let $\|\cdot\|$ be a lattice-valued interpretation satisfying $(\vec{\epsilon} - \top)$ and let d be a semi-pseudometric on $\text{target}(\|\cdot\|)$. The following are equivalent:*

1. d satisfies (strong \wedge rule) and the triangle inequality.
2. d is the trivial metric on $\text{target}(\|\cdot\|)$.

The above facts follow from the equality case in (strong \wedge rule): for any sequences $\vec{\varphi}, \vec{\psi}_1, \vec{\psi}_2$, if $\vec{\psi}_1 \wedge \vec{\varphi} = \vec{\psi}_2 \wedge \vec{\varphi}$ then $\vec{\psi}_1$ and $\vec{\psi}_2$ have to be equidistant from $\vec{\varphi}$. Removing this assumption yields a weakening of (strong \wedge rule) which no longer support the trivialisation result above:

If $\|\vec{\varphi}_1\| \wedge \|\vec{\varphi}_2\| \leq \|\vec{\varphi}_1\| \wedge \|\vec{\varphi}_3\|$ then $\delta(\vec{\varphi}_1, \vec{\varphi}_2) \geq \delta(\vec{\varphi}_1, \vec{\varphi}_3)$ (weak \wedge rule)

Fact 8.4. d_H^{ham} , d_\ominus and d_α do not satisfy (weak \wedge rule).

Fact 8.5. For any set-valued $\|\cdot\|$, d_C verifies (weak \wedge rule).

Corollary 8.6. $\|\cdot\|^t$ is an intersective interpretation which yield a d_C that verifies (weak \wedge rule), the triangle inequality and is not trivial.

8.2 Disjunction Axioms

Assume that $\|\cdot\|$ is lattice-valued. The following axiom is very mild.

$$\text{If } \|\vec{\psi}\| = \|\vec{\varphi}_1\| \vee \|\vec{\varphi}_2\| \text{ then } d(\vec{\varphi}_1, \vec{\psi}) \leq d(\vec{\varphi}_1, \vec{\varphi}_2) \quad (\vee \text{ rule})$$

Fact 8.7. Assume that $\text{target}(\|\cdot\|) = \wp(W)$ for some non-empty W and that $\|\cdot\|$ is lattice-valued. Let δ be a metric on W . d_H^δ satisfies $(\vee \text{ rule})$.

Fact 8.8. d_\ominus, d_α and d_C also satisfy $(\vee \text{ rule})$.

9 Future Directions

So far we have focused on isolating an abstract concept of semanticity for metrics. We have explored general axioms that help us express components of this concept. We have also identified more specific axioms that were candidates at defining the contour of a notion of ‘good behavior’ for semantic metrics, and thus at being criterion for evaluating such metrics. We have done this at an abstract level, considering conversations as sequences of formulas where one conversational agent plays a sequence of formulas after the other. The conversation thus has the structure of a (syntactic) monoid with a syntactic composition operation of concatenation. Corresponding to sequences of formulas is their abstract interpretation in a different, semantic space; the generic notion of a semantic interpretation furnishes the correspondence, mapping these sequences into the semantic space.

Coming back to the goals outlined in the introduction of this paper, the next step of our work is to extend this perspective to structures that represent real conversations. We mention a few directions here, each of which can be explored independently. In order to do this, we need to fill in this abstract framework with notions that capture aspects of conversational content at various levels of detail. A first step is to refine the notion of sequence of formulas into something that preserves more of the logical form of conversations. Most models of discourse interpretation assume a more structured representation of conversations, e.g., trees or graphs, in which elementary discourse units are linked together via discourse relations to form more complex discourse units. Using such structures to represent conversations would require us to adapt the structural properties of interpretation functions considered in Sect. 3.2 to be able to reflect the semantics of discourse relations and the units they link together. Second we would need to revisit the axioms that make use of concatenation, replacing the latter with a notion of a graph update or graph extension.

Furthermore, to deal adequately with some natural language phenomena such as questions, commands, agreements and disagreements among speakers, explicit or implicit corrections, it is natural to assume additional structure for semantic spaces, on top of that provided by general lattices. This additional semantic structure could also serve to refine some of our axioms, in particular those making hypotheses on lattice-theoretic relations between the semantic interpretations of two conversations.

Different notions of semantic interpretation carry different amounts of the initial syntactic structure into the semantic space. The classical notion of information content for a discourse erases all structural information, mapping discourses such as *Jane fell because John pushed her* and *John pushed Jane so she fell* into the exact same semantic interpretation (either a set of possible worlds or a set of world assignment pairs as in SDRT and other dynamic semantic theories). Differently structured discourses, even when they share the same meaning, however, may exhibit different semantic and pragmatic behavior, concerning the possibility of future coreferences and of ways to extend the conversation. Intuitions dictate that these features are important for a notion of conversational similarity. It will therefore be important to test metrics defined on more structurally-conservative spaces, for instance conserving some aspects of the conversational graph. These metrics should match intuitions as to how far two real conversations are from each other.

10 Conclusions

Our first task was to explore the concept of a semantic metric by identifying a certain number of reasonable axioms that characterize the idea of ‘semanticity’ for a distance. We clarified the relation between these different axioms and the triangle inequality, and we mapped out a lattice of axioms in terms of their logical strength. Next, we explored a structured list of candidate axioms or desirable properties for any semantic metric. We found several to be too demanding, in the sense that under some structural constraints on the interpretation function and on the distance, they could only be satisfied by the trivial metric. These axioms divide into a certain number of categories. First, we considered a certain number of axioms pertaining to general properties of semantic metrics, including arguably mild assumptions about their structure, their domain and their insensitivity to the choice of signature. Then, we considered preservation axioms that carry a general idea of coherence between the relative proximity of sequences and of their extensions. Finally we considered axioms that are more specific to a lattice- or a set-theoretic approach.

We concentrated on the foundational case of conversations as sequences of propositional formulae with a classical truth functional interpretation by studying four semantically induced metrics that looked intuitively promising (based respectively on the ideas of symmetric semantic difference, semantic proportionality, Hausdorff metric and on possible continuations). We now have a clear picture of their different behavior. Overall however, these metrics satisfy only few of our axioms that do not lead to a triviality result. One reason for this are the very strong structural hypotheses behind the set-theoretic, classical interpretation of the language of the propositional calculus. A further exploration of these axioms in the context of interpretation into structures like lattices with fewer structural hypotheses and of more general families of metrics remains to be done. We hope that the abstract setting that we have set up in this paper can serve a first step towards achieving this goal.

A Appendix: Selected Proofs

(PROOF OF FACT 5.2). Assume that for every $\vec{\chi}_1, \vec{\chi}_2$ with $\|\vec{\chi}_1\| = \|\vec{\chi}_2\|$ we have $d(\vec{\varphi}, \vec{\chi}_1) = d(\vec{\psi}, \vec{\chi}_2)$ (i). Now take some $\vec{\chi}$. We have $\|\vec{\chi}\| = \|\vec{\chi}\|$. Hence by (i), we have $d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$. But $\vec{\chi}$ was arbitrary, hence $\forall \chi \quad d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$. It follows, by (sem separation), that $\|\vec{\varphi}\| = \|\vec{\psi}\|$. QED

(PROOF OF FACT 5.3). Assume that $\forall \vec{\chi}, d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$. In particular $d(\vec{\varphi}, \vec{\psi}) = d(\vec{\psi}, \vec{\psi}) = 0$. Hence by (zero \Rightarrow sem \equiv), $\|\vec{\varphi}\| = \|\vec{\psi}\|$. QED

(PROOF OF FACT 5.4). The right to left direction follows from Fact 5.3. For the left to right direction, assume that $d(\vec{\varphi}, \vec{\psi}) = 0$ (i). Take any $\vec{\chi}$. By triangle inequality, $d(\vec{\chi}, \vec{\varphi}) \leq d(\vec{\chi}, \vec{\psi}) + d(\vec{\psi}, \vec{\varphi})$. Hence, by (i) we have $d(\vec{\chi}, \vec{\varphi}) \leq d(\vec{\chi}, \vec{\psi})$. Similarly we have $d(\vec{\chi}, \vec{\psi}) \leq d(\vec{\chi}, \vec{\varphi})$. Hence $d(\vec{\chi}, \vec{\psi}) = d(\vec{\chi}, \vec{\varphi})$. But χ was arbitrary, hence for all χ we have $d(\vec{\chi}, \vec{\psi}) = d(\vec{\chi}, \vec{\varphi})$. By (sem separation), it follows that $\|\vec{\varphi}\| = \|\vec{\psi}\|$. QED

(PROOF OF FACT 5.7). $\|p \neg p\| = \|q \neg q\|$ but $\delta_{count}(p \neg p, q \neg q) = 4$. QED

(PROOF OF FACT 5.8). $\|p \neg p\| = \|q \neg q\|$ but $\delta_{synt, count}(p \neg p, q \neg q) = 6$. QED

(PROOF OF FACT 5.9). Take some φ such that $\text{card}(\|\varphi\|) \geq 2$. QED

(PROOF OF FACT 5.10). Assume that $\|\vec{\varphi}\| = \|\vec{\psi}\|$. By (sem preservation) we have $d(\vec{\varphi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$. In particular we have $d(\vec{\varphi}, \vec{\psi}) = d(\vec{\psi}, \vec{\psi}) = 0$ QED

(PROOF OF FACT 5.11). Assume that $\|\vec{\varphi}_1\| = \|\vec{\psi}_1\|$. Take some $\vec{\chi}$. We have $\|\vec{\chi}\| = \|\vec{\chi}\|$. Hence by (sem induced) we have $d(\vec{\varphi}_1, \vec{\chi}) = d(\vec{\psi}_1, \vec{\chi})$. QED

(PROOF OF FACT 5.12). Assume that $\|\vec{\varphi}_1\| = \|\vec{\psi}_1\|$ (i) and $\|\vec{\varphi}_2\| = \|\vec{\psi}_2\|$ (ii). By triangle inequality we have:

$$\begin{aligned} d(\vec{\varphi}_1, \vec{\varphi}_2) &\leq \underbrace{d(\vec{\varphi}_1, \vec{\psi}_1)}_{0, \text{ by (sem} \equiv \Rightarrow \text{ zero)}} + d(\vec{\psi}_1, \vec{\varphi}_2) \\ d(\vec{\psi}_1, \vec{\varphi}_2) &\leq d(\vec{\psi}_1, \vec{\psi}_2) + \underbrace{d(\vec{\psi}_2, \vec{\varphi}_2)}_{0, \text{ by (sem} \equiv \Rightarrow \text{ zero)}} \end{aligned}$$

Hence, $d(\vec{\varphi}_1, \vec{\varphi}_2) \leq d(\vec{\psi}_1, \vec{\psi}_2)$. Similarly, we have $d(\vec{\varphi}_1, \vec{\varphi}_2) \geq d(\vec{\psi}_1, \vec{\psi}_2)$. QED

(PROOF OF COROLLARY 5.13). By Fact 5.12, d satisfies (sem induced), hence for every $\vec{\varphi}_1, \vec{\varphi}_2, \vec{\psi}_1, \vec{\psi}_2$ with $\|\vec{\varphi}_1\| = \|\vec{\psi}_1\|$ and $\|\vec{\varphi}_2\| = \|\vec{\psi}_2\|$ we have

$$d(\vec{\varphi}_1, \vec{\varphi}_2) = d(\vec{\psi}_1, \vec{\psi}_2)$$

It follows that $d(\|\vec{\varphi}\|, \|\vec{\psi}\|) := d(\vec{\varphi}, \vec{\psi})$ is well-defined. Moreover for any $\vec{\varphi}$, $d(\|\vec{\varphi}\|, \|\vec{\varphi}\|) = d(\vec{\varphi}, \vec{\varphi}) = 0$. Triangle inequality is proven similarly. QED

(PROOF OF FACT 5.14). First observe, that by triangle inequality, we have

$$d(\vec{\chi}, \vec{\varphi}) \leq d(\vec{\chi}, \vec{\psi}) + d(\vec{\psi}, \vec{\varphi})$$

Now, assume that $\vec{\varphi} \equiv \vec{\psi}$. By (sem $\equiv \Rightarrow$ zero) we have $d(\vec{\psi}, \vec{\varphi}) = 0$, hence $d(\vec{\chi}, \vec{\varphi}) \leq d(\vec{\chi}, \vec{\psi})$. Similarly, $d(\vec{\chi}, \vec{\psi}) \leq d(\vec{\chi}, \vec{\varphi})$ which proves (1). QED

(PROOF OF FACT 5.15). Assume that $\forall \chi$ we have $d(\varphi, \chi) = d(\psi, \chi)$ (i). Take some $\vec{\chi}_1, \vec{\chi}_2$ with $\|\vec{\chi}_1\| = \|\vec{\chi}_2\|$. By (sem $\equiv \Rightarrow$ zero) we have $d(\vec{\chi}_1, \vec{\chi}_2) = 0$ (ii). By triangle inequality we have:

$$\begin{aligned} d(\vec{\varphi}, \vec{\chi}_1) &\leq d(\vec{\varphi}, \vec{\chi}_2) + \underbrace{d(\vec{\chi}_2, \vec{\chi}_1)}_{0, \text{ by (ii)}} \\ d(\vec{\varphi}, \vec{\chi}_2) &\leq d(\vec{\varphi}, \vec{\chi}_1) + \underbrace{d(\vec{\chi}_1, \vec{\chi}_2)}_{0, \text{ by (ii)}} \end{aligned}$$

Hence $d(\vec{\varphi}, \vec{\chi}_1) = d(\vec{\varphi}, \vec{\chi}_2)$. Moreover by (i) we have $d(\vec{\varphi}, \vec{\chi}_2) = d(\vec{\psi}, \vec{\chi}_2)$. Hence $d(\vec{\varphi}, \vec{\chi}_1) = d(\vec{\psi}, \vec{\chi}_2)$. Since $\vec{\chi}_1, \vec{\chi}_2$ were arbitrary, it follows by (min sem separation), that $\|\vec{\varphi}\| = \|\vec{\psi}\|$. QED

(PROOF OF FACT 6.12). Let v be an isotone upper valuation and $z \leq x, y$. Assume without loss of generality that $v(y) \geq v(x)$. We can write, for $0 \leq \alpha \leq 1$:

$$-\alpha \cdot 2^{-v(y)} \geq -2^{-v(y)} \geq -2^{-v(x)},$$

which ensures that

$$-\alpha \cdot 2^{-v(y)} - \frac{1}{\alpha} 2^{-v(x)} \leq -2^{-v(y)} - 2^{-v(x)}.$$

Instantiating this result for $\alpha = 2^{v(z)-v(x)}$ yields after development

$$-2^{v(z)-v(x)-v(y)} - 2^{-v(z)} \leq -2^{-v(y)} - 2^{-v(x)}.$$

Since v is an upper-valuation, we have $-v(x \vee y) \geq v(z) - v(x) - v(y)$ and thus

$$-2^{-v(x \vee y)} - 2^{-v(z)} \leq -2^{v(z)-v(x)-v(y)} - 2^{-v(z)} \leq -2^{-v(y)} - 2^{-v(x)}$$

i.e., $w(x \vee y) + w(z) \leq w(x) + w(y)$ which concludes the proof.

The case $v(x) \geq v(y)$ is symmetrically dealt with. QED

(PROOF OF FACT 6.16). Let $\vec{\varphi} := (p_1 \wedge p_2) \vee (\neg p_1 \wedge \neg p_2 \wedge \neg p_3)$ and $\vec{\psi} := (p_1 \wedge (p_2 \rightarrow p_3))$, and assume some intersective interpretation of concatenation. We have $\|\vec{\varphi} \vec{\psi}\| = \|p_1 \wedge p_2 \wedge p_3\|$. $d_H^{ham}(\vec{\varphi}, \vec{\varphi} \vec{\psi}) = 3$, but $d_H^{ham}(\vec{\varphi}, \vec{\psi}) = 1$. QED

(PROOF OF FACT 6.18). For any φ , d_H^{ham} is neither well-defined for (φ, \perp) nor for (\perp, φ) . QED

(PROOF OF FACT 6.21). Adding a new propositional letter that does not occur in either sequence will not affect the symmetric difference of the range of formulas, nor the symmetric difference of the respective signature. Allowing for the negation of the propositional letter that was previously forbidden will not change the sets either. QED

(PROOF OF FACT 7.1). Assume that $d(\varphi, \chi) \leq d(\psi, \chi)$ then $d(\varphi\psi\chi, \chi\psi\chi) \leq d(\psi\psi\chi, \chi\psi\chi)$. By exchange, contraction, expansion, and (sem $\equiv \Rightarrow$ zero), we have $d(\psi\psi\chi, \chi\psi\chi) = 0$. Hence $d(\varphi\psi\chi, \chi\psi\chi) = 0$. By exchange, contraction, expansion, and (sem $\equiv \Rightarrow$ zero), we have $d(\varphi\psi\chi, \chi\psi) = 0$. Concluding our proof. QED

(PROOF OF COROLLARY 7.2). We only give the idea of the proof. The idea of the proof is to define a linear order on \mathbf{L}^* compatible with \leq_o . By induction, using Fact 7.1 we first show the claim for formulas in the same o -equivalence class, then we show that the claim propagate downward, that is for every $\vec{\psi} \in \downarrow[\vec{\varphi}]$. Finally we show that the claim propagates with transitive closure. QED

(PROOF OF COROLLARY 7.3). Direct from Fact 5.15 and Corollary 7.2. QED

(PROOF OF FACT 7.4). Let $k \geq 2$, $n = 2k$. Now let

$$\begin{aligned}\varphi &:= p_1 \rightarrow (\neg p_2 \wedge \dots \wedge \neg p_n) \wedge \neg p_1 \rightarrow (p_2 \wedge \dots \wedge p_n), \\ \psi &:= p_1 \wedge \neg p_2 \wedge \dots \wedge p_{2k-1} \wedge \neg p_{2k},\end{aligned}$$

$\chi := p_1 \wedge \dots \wedge p_n$ and $\varphi_0 := p_1$. Since $k \geq 2$ we have

$$\begin{aligned}1 &= d_H^{ham}(\varphi, \chi) < d_H^{ham}(\varphi, \chi) = k, \text{ and,} \\ n - 1 &= d_H^{ham}(\varphi\varphi_0, \chi\varphi_0) > d_H^{ham}(\varphi\varphi_0, \chi\varphi_0) = k = n/2\end{aligned}$$

Concluding our proof. QED

(PROOF OF FACT 7.6). (1 \Rightarrow 2). Take some $\vec{\varphi}, \vec{\psi}$. By contraction, expansion and exchange we have

$$\|\vec{\varphi}\vec{\varphi}\vec{\psi}\| = \|\vec{\psi}\vec{\varphi}\vec{\psi}\|$$

Hence by (sem $\equiv \Rightarrow$ zero), $d(\vec{\varphi}\vec{\varphi}\vec{\psi}, \vec{\psi}\vec{\varphi}\vec{\psi}) = 0$ (i). Hence $d(\vec{\varphi}\vec{\varphi}\vec{\psi}, \vec{\psi}\vec{\varphi}\vec{\psi}) \leq d(\vec{\psi}\vec{\varphi}\vec{\psi}, \vec{\psi}\vec{\varphi}\vec{\psi})$ (i). Now, let $\vec{\chi} = \vec{\psi}$ and $\vec{\varphi}_0 = \vec{\varphi}\vec{\psi}$. By (i) and (uniform anti-preservation) we have

$$\text{If } d(\vec{\varphi}\vec{\varphi}\vec{\psi}, \vec{\psi}\vec{\varphi}\vec{\psi}) \leq d(\vec{\psi}\vec{\varphi}\vec{\psi}, \vec{\psi}\vec{\varphi}\vec{\psi}) \text{ then } d(\vec{\varphi}, \vec{\psi}) \leq d(\vec{\psi}, \vec{\psi}) = 0$$

Hence by (i), $d(\vec{\varphi}, \vec{\psi}) = 0$. Concluding the proof for this direction. The other direction is trivial. QED

(PROOF OF FACT 7.8). Let $n \in \omega$ be such that $n > 5$. Moreover let: $\varphi := (p_3 \wedge \dots \wedge p_n) \wedge \neg(p_1 \wedge p_2)$, $\psi_1 := (p_1 \vee \dots \vee p_n) \wedge \neg(p_2 \wedge \dots \wedge p_n)$ and $\psi_2 := (p_2 \wedge \dots \wedge p_n)$. We have $d_\alpha(\varphi, \psi_1) = 1 - \frac{2}{n+3} = \frac{n+1}{n+3} > d_\alpha(\varphi, \psi_2) = 1 - \frac{1}{4} = \frac{3}{4}$. But we have $d_\alpha(\varphi, \varphi\psi_1) = 1 - \frac{2}{3} = \frac{1}{3} < d_\alpha(\varphi, \varphi\psi_2) = 1 - \frac{1}{3} = \frac{2}{3}$. QED

(PROOF OF FACT 7.9). Take some $\vec{\varphi}, \vec{\psi}$. By (coherent deviation) we have

$$\text{If } d(\vec{\varphi}, \vec{\epsilon}) < d(\vec{\psi}, \vec{\epsilon}) \text{ then } d(\vec{\varphi}, \vec{\varphi}\vec{\epsilon}) < d(\vec{\psi}, \vec{\psi}\vec{\epsilon})$$

$$\text{If } d(\vec{\psi}, \vec{\epsilon}) < d(\vec{\varphi}, \vec{\epsilon}) \text{ then } d(\vec{\psi}, \vec{\psi}\vec{\epsilon}) < d(\vec{\varphi}, \vec{\varphi}\vec{\epsilon})$$

Since $\vec{\varphi}, \vec{\psi}$ were arbitrary, it follows that for any $\vec{\varphi}, \vec{\psi}$, $d(\vec{\varphi}, \vec{\epsilon}) = d(\vec{\psi}, \vec{\epsilon})$. In particular $d(\vec{\varphi}, \vec{\epsilon}) = d(\vec{\psi}, \vec{\epsilon}) = d(\vec{\epsilon}, \vec{\epsilon}) = 0$. Hence by (triangle inequality) we have $\forall \vec{\varphi}, \vec{\psi} d(\vec{\varphi}, \vec{\psi}) \leq d(\vec{\varphi}, \vec{\epsilon}) + d(\vec{\epsilon}, \vec{\psi}) = 0$. QED

(PROOF OF FACT 7.10). (1 \Rightarrow 2). Take some $\vec{\varphi}, \vec{\psi}$. We have

$$0 = d(\vec{\varphi}, \vec{\varphi}\vec{\epsilon}) \leq d(\vec{\varphi}, \vec{\varphi}\vec{\psi})$$

Hence by (converse strong action pref) $d(\vec{\varphi}, \vec{\epsilon}) \leq d(\vec{\varphi}, \vec{\psi})$. But ψ was arbitrary, hence, in particular $d(\vec{\varphi}, \vec{\epsilon}) \leq d(\vec{\varphi}, \vec{\varphi}) = 0$. But φ was arbitrary as well, hence $\forall \chi d(\vec{\varphi}, \vec{\epsilon}) = 0$. Hence by triangle inequality for any formula $\vec{\varphi}, \vec{\psi}$ we have $d(\vec{\varphi}, \vec{\psi}) \leq d(\vec{\varphi}, \vec{\epsilon}) + d(\vec{\epsilon}, \vec{\psi}) = 0$. Concluding our proof. QED

(PROOF OF FACT 7.11). (1 \Rightarrow 2). Take some $\vec{\varphi}$ and $\vec{\psi}$. We have

$$0 = d(\vec{\varphi}, \vec{\varphi}\vec{\epsilon}) \leq d(\vec{\epsilon}, \vec{\epsilon}\vec{\epsilon}) = 0$$

By (converse coherent deviation) $d(\vec{\varphi}, \vec{\epsilon}) \leq d(\vec{\epsilon}, \vec{\epsilon}) = 0$ (i). Similarly, we have $d(\vec{\psi}, \vec{\epsilon}) = 0$ (ii). By (i), (ii) and triangle inequality we have $d(\vec{\varphi}, \vec{\psi}) \leq d(\vec{\varphi}, \vec{\epsilon}) + d(\vec{\epsilon}, \vec{\psi}) = 0$. Concluding our proof. QED

(PROOF OF FACT 8.1). Take some $\vec{\chi}$ and assume that $\|\vec{\varphi}\| \leq \|\vec{\psi}\|$. Since $\text{target}(\|\cdot\|)$ is a lattice. We have $\|\vec{\chi}\| \wedge \|\vec{\varphi}\| \leq \|\vec{\chi}\| \wedge \|\vec{\psi}\|$. Hence by (strong \wedge rule) $\delta(\vec{\chi}, \vec{\varphi}) \geq \delta(\vec{\chi}, \vec{\psi})$. QED

(PROOF OF COROLLARY 8.2). Assume that $\|\vec{\varphi}\| \leq \|\vec{\psi}\|$. Since d satisfies (strong \wedge rule), we have by Fact 8.1 we have in particular $0 = \delta(\vec{\varphi}, \vec{\varphi}) \geq \delta(\vec{\varphi}, \vec{\psi}) = \delta(\vec{\psi}, \vec{\varphi})$. QED

(PROOF OF COROLLARY 8.3). (1 \Rightarrow 2). Take two arbitrary $\vec{\varphi}, \vec{\psi}$. By $(\vec{\epsilon} - \top)$ we have $\|\vec{\varphi}\| \leq \|\vec{\epsilon}\|$ and $\|\vec{\psi}\| \leq \|\vec{\epsilon}\|$. Since d satisfies (strong \wedge rule), it follows by Corollary 8.2 that $d(\vec{\varphi}, \vec{\epsilon}) = d(\vec{\epsilon}, \vec{\psi}) = 0$. By triangle inequality it follows that $d(\vec{\varphi}, \vec{\psi}) = 0$. The (2 \Rightarrow 1) direction is trivial. QED

(PROOF OF FACT 8.5). Take $\vec{\varphi}, \vec{\psi}_1, \vec{\psi}_2$ with $\vec{\varphi} \cap \vec{\psi}_1 \subsetneq \vec{\varphi} \cap \vec{\psi}_2$. And consider the cardinalities assigned in Fig. 2. Let $X_1 := \alpha_1 + \eta_0$ and let $X_2 := \alpha_2 + \eta_0$. Using these cardinalities and inserting them in the expression of the distance, gives us:

$$\begin{aligned} d_C(\varphi, \psi_1) &= \frac{2^{\text{card}(\|\vec{\varphi}\|)} + 2^{\text{card}(\|\vec{\psi}_1\|)} - 2 \cdot 2^{\text{card}(\|\vec{\varphi}\| \cap \|\vec{\psi}_1\|)}}{2^{\text{card}(\|\vec{\varphi}\|)} + 2^{\text{card}(\|\vec{\psi}_1\|)}} \\ &= \frac{2^{X_\varphi + \eta_1 + \eta_2} + 2^{X_1 + \eta_1} - 2 \cdot 2^{\eta_1}}{2^{X_\varphi + X_1 + 2\eta_1 + \eta_2}} \end{aligned}$$

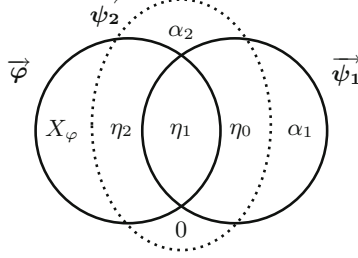


Fig. 2. Assigning cardinalities to the respective intersections.

Similarly:

$$d_C(\varphi, \psi_2) = \frac{2^{X_\varphi + \eta_1 + \eta_2} + 2^{X_2 + \eta_1 + \eta_2} - 2 \cdot 2^{\eta_1 + \eta_2}}{2^{X_\varphi + X_2 + 2\eta_1 + 2\eta_2}}$$

From the two previous expression, after simplifications we find:

$$d_C(\varphi, \psi_1) - d_C(\varphi, \psi_2) = \frac{2^{X_\varphi} (2^{X_2 + \eta_2} - 2^{X_1}) + 2(2^{X_1} - 2^{X_2})}{2^{X_\varphi + X_1 + X_2 + \eta_1 + \eta_2}}$$

From the assumption that $\vec{\varphi} \cap \vec{\psi}_1 \subsetneq \vec{\varphi} \cap \vec{\psi}_2$, it follows that $\eta_2 \geq 1$, hence:

$$2^{X_\varphi} (2^{X_2 + \eta_2} - 2^{X_1}) + 2(2^{X_1} - 2^{X_2}) \geq 2^{X_2 + \eta_2} - 2^{X_2 + 1} + 2^{X_1 + 1} - 2^{X_1} \geq 0$$

which concludes the proof. QED

(PROOF OF FACT 8.7). Take $\vec{\varphi}_1, \vec{\varphi}_2$. Assume that $\|\vec{\psi}\| = \|\vec{\varphi}_1\| \vee \|\vec{\varphi}_2\|$ (i). By definition, we have

$$d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2) = \max\left\{ \max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\varphi}_2\|} \delta(x, y), \max_{y \in \|\vec{\varphi}_2\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) \right\}$$

Hence, we are in one of two cases.

- (1) $\max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\varphi}_2\|} \delta(x, y) = d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$, or,
- (2) $\max_{y \in \|\vec{\varphi}_2\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) = d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$.

Case 1. There are two subcases.

Subcase 1a. Assume that $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = \max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\psi}\|} \delta(x, y)$ (a). By (i),

$$\max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\psi}\|} \delta(x, y) \leq \max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\varphi}_2\|} \delta(x, y)$$

By (a) and (1) we have $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$.

Subcase 1b. $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = \max_{y \in \|\vec{\psi}\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y)$ (b). Since δ is a metric we have $\max_{y \in \|\vec{\varphi}_1\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) = 0$. Hence by (i) and (b) we have

$d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = \max_{y \in \|\vec{\varphi}_2\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y)$ (ii). But by (1) and definition of $d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$ we have $\max_{y \in \|\vec{\varphi}_2\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$ (iii). By (ii) and (iii) we have $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$.

Case 2. Since δ is a metric we have

$$\max_{y \in \|\vec{\varphi}_1\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) = 0$$

Hence by (2), we have

$$\begin{aligned} \max_{y \in \|\vec{\psi}\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) &= d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2) \\ &= \max_{y \in \|\vec{\varphi}_2\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y) \\ &= d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2) \end{aligned} \tag{iv}$$

We now consider two subcases.

Subcase 2a. $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = \max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\psi}\|} \delta(x, y)$ (a). But since δ is a metric we have $\max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\psi}\|} \delta(x, y) \leq \max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\varphi}_2\|} \delta(x, y)$ (v). But by definition $d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$ and (2) we have $\max_{x \in \|\vec{\varphi}_1\|} \min_{y \in \|\vec{\varphi}_2\|} \delta(x, y) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$ (vi). By (v), (vi) and (a), it follows that $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$.

Subcase 2b. Assume that. $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = \max_{y \in \|\vec{\psi}\|} \min_{x \in \|\vec{\varphi}_1\|} \delta(x, y)$ (b). By (iv), it follows that $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) = d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$.

Hence in all cases $d_H^\delta(\vec{\varphi}_1, \vec{\psi}) \leq d_H^\delta(\vec{\varphi}_1, \vec{\varphi}_2)$. Concluding our proof. QED

References

- Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: partial meet contraction and revision functions. *J. Symbolic Logic* **50**(2), 510–530 (1985)
- Asher, N., Lascarides, A.: *Logics of Conversation*. Cambridge University Press, Cambridge (2003)
- Black, E., Abney, S.P., Flickenger, D., Gdaniec, C., Grishman, R., Harrison, P., Hindle, D., Ingria, R., Jelinek, F., Klavans, J.L., Liberman, M., Marcus, M.P., Roukos, S., Santorini, B., Strzalkowski, T.: A procedure for quantitatively comparing the syntactic coverage of english grammars. In: *Proceedings of a Workshop held at Pacific Grove, Speech and Natural Language, California, USA, 19–22 February 1991*, Morgan Kaufmann (1991)
- Eiter, T., Mannila, H.: Distance measures for point sets and their computation. *Acta Informatica* **34**(2), 109–133 (1997)
- Grice, H.P.: *Logic and conversation*. Studies in the Way of Words, pp. 22–40. Harvard University Press, Cambridge, MA (1967)
- Lehmann, D., Magidor, M., Schlechta, K.: Distance semantics for belief revision. *J. Symbolic Logic* **66**(1), 295–317 (2001)
- Monjardet, B.: Metrics on partially ordered sets - a survey. *Discrete Math.* **35**(13), 173–184 (1981). Special Volume on Ordered Sets

- Popper, K.R.: Conjectures and Refutations: The Growth of Scientific Knowledge. Harper and Row, New York (1968)
- Sampson, G.: A proposal for improving the measurement of parse accuracy. *Int. J. Corpus Linguist.* **5**, 53–68 (2000)